# The increasing availability of official datasets: methods, limitations and opportunities for studies of education

Stephen Gorard
University of Birmingham
s.gorard@bham.ac.uk

## Abstract

The re-use of existing and official data has a very long and largely honourable history in education and social science. The principal change in the sixty years since the first issue of the British Journal of Educational Studies has been the increasing range, availability, and quality of existing numeric datasets. New and valuable fields of endeavour have formed around specific applications of these datasets, and some promising analytical techniques have been devised to deal with them. At the same time, the opportunities provided by these datasets have thrown up fascinating methodological and other challenges. This paper presents a brief summary of all of these developments, with illustrative examples. The paper then considers two fields of endeavour that have been particularly valuable in education studies – the political arithmetic tradition of origins and destinations, and the school effectiveness and improvement. Both have used similar datasets but for different purposes and so reaching rather different conclusions, with considerably different take-up by policy-makers. The paper ends by envisaging some of the ways in which these and other fields using secondary data may develop in the near future.

## Introduction

Research evidence is at the heart of finding out more about the social and education world. Creating new and primary evidence can be costly and inefficient, while using evidence at third-hand, as in a review of literature, can be distant and distorting. The latter is 'third-hand' because the analyst is presenting a summary of what a previous author presented about an analysis of data (Hakim 1982). A useful compromise is to use existing data, where it is available, for a new analysis. In any area of research in education, life long and society wide, large relevant numeric datasets will already exist. These datasets are likely to be larger in scope and scale, and higher quality in terms of completeness and validity, than anything a researcher could generate through primary fieldwork. And they can be accessed directly, sorted for a new purpose, and combined with other sources of data. They can often be analysed very simply because they avoid the need for traditional statistical generalisation. This paper looks at the trend towards greater use of such secondary data, with its main focus on numeric datasets and the UK. However, several of the points made will be applicable to any country, and to other forms of secondary data (such as video archives).

Secondary statistics as used in social science derive from the idea of political arithmetic in the 1660s in the UK (Porter 1986). The purpose of this political arithmetic was to promote sound, well-informed state policy, and so to raise life expectancy and population figures, and reform health, education and the handling of

crime. It largely pursued these aims by laying bare issues of deprivation, inequality, and the stratification of opportunities. In this way it encouraged governments, charities and other activists to tackle these problems, either by informing them more fully or by shaming them through widespread publication. More recent calls to make better use of existing records in social science have come from Bulmer (1980), and have been strongly encouraged by a succession of commentators in education since (Heath 2000, Gorard 2001, Smith 2008).

This paper re-presents a summary of the reasons why the use of secondary is growing, describes some important sources of data, some limitations, and some of the research design and methods issues that arise. The paper continues with a description of two rather different traditions of using secondary data, both highly influential in education studies, and concludes with some brief suggestions of how things might develop in the near future.

## Why use secondary data?

The cumulation of evidence is important in social science, and so the consideration of existing data is essential for at least the early part of an empirical investigation - if only to ensure that any new data proposed does not already exist. Funding bodies allocating publicly-funded grants or commissioning research, such as the UK Economic and Social Research Council, require applicants to show that they have looked to see if the data they require already exists, and to present evidence that it does not. In addition, once a publicly-funded project is completed the datasets generated must be lodged with a public data archive, thereby increasing the chance for each new proposal that something similar already exists in the archive. Cumulation also includes replication and re-analysis, which is a vital part of keeping analysts 'honest'.[1]

Sometimes, the use of existing records is inevitable. As illustrated in the earliest issues of BJES, when potential respondents no longer exist, as would be the case for someone who was now researching nineteenth century schooling, recall and recollection are no longer possible. For historians of education, the use of pre-existing data is therefore natural and useful. Otherwise, speed and cost are probably the most obvious advantages of using secondary data. Since the data already exists it is generally quicker to 'collect', involving less travel and minimal cost. This means that the researcher can make more progress in a given time period. Many valuable datasets are available free of charge. Even where there is a cost for access to a dataset this is likely to end up cheaper than incurring the costs of travel, telephone, printing, postage, and subsistence involved in carrying out primary data collection.

[1] For an important example, see the secondary analysis of part of the data used in the lavishly expensive teacher effectiveness study by Hay McBer study (Gorard 2002). In trying to relate reported teacher quality with the value-added scores of their pupils, Hay McBer found a correlation of +0.43 using a heavily selected sample of 128 teachers. But this association was not considered high enough, and so the report dropped 32 teachers who had excellent pupil progress scores but who were not considered good teachers or who were considered good teachers but had poor pupil progress. The new 'sample' of 96 teachers yielded a correlation of +0.79. This was the headline figure published. Such deception and/or incompetence has to be exposed by reconsideration or replication using the same dataset (see also Gorard 2008).

For example, in assembling the data for my early work on the socio-economic composition of schools (Gorard and Fitz 1998), I needed the annual census returns for schools from Local Education Authorities (LEAs) for as many previous years as available. These records were only held centrally by the Welsh Office (as it was then) for the prior two years. To get earlier records I had to negotiate access to each LEA, and in travel to their offices, spending half a day in a dusty cupboard full of the school census archives. Even so, I completed the study for a total sum of less than £100 for travel, postage and telephone. This £100 project, while still the subject of some debate, went on to change the field of school choice research and attracted both media and political interest on an international scale.

Secondary data might be used to help select the sample for a further in-depth study. It can provide the appropriate figures for each strata in a stratified sample. It can be used to assess the quality of an achieved sample by providing some background figures for the population. These figures can then be used to re-weight the sample if there is clear bias in its composition. For example, Gorard and Rees (2002) used the electoral register in each region to select cases for their door-to-door household study of patterns of lifelong learning. The ten-yearly census of population is also a useful way of characterising the population of different regions, and so selecting an areal sample (Gorard and Selwyn 2005, Selwyn et al. 2006). The annual schools census in England can be used to select schools to represent the range of pupil intakes (Gorard et al. 2003). In all of these ways, the large-scale dataset can be argued to yield cases for more detailed study that represent the larger picture. This approach overcomes some of the deficiencies of both kinds of data, providing an audit trail of generalisability for in-depth case studies perhaps (Gorard with Taylor 2004). With increasing access to large-scale datasets researchers are now strongly encouraged by funders to preface their studies routinely with an analysis of the relevant population figures (Rendall 2003), before moving on to work with in-depth data or case studies.

Contextual secondary data can also be used to show that a problem exists that needs to be addressed using other techniques, and to begin to describe the nature of that problem. When investigating the causes of increasing crime in city centres, for example, the researcher needs to show via secondary data that this problem does actually exist. Many 'moral panics' are based on misreading of the existing data. Secondary data can also show something about the nature of the problem being investigated. Is the increase in all categories of crime, and is it manifested differently in different cities? It only makes sense to move to the primary phase of an investigation once the study has been justified, the sample has been created, and the existing data examined for patterns - all via secondary data.

For example, in the late 1990s in the UK there was considerable concern about the apparent underachievement of boys. A lot of research focussed on why boys were failing, and why girls were, for the first time, ahead of boys in terms of examination and assessment. But this research appears to have been looking at the wrong research questions in a number of ways (Gorard et al. 2001). The difference between boys and girls, where it appears, is not in terms of failing. It is, or at least was, at the highest levels of attainment, such as grade A at A-level (Table 1). In subjects like maths and sciences, there was actually very little difference between the results of boys and girls. There used to be a gap in favour of boys at the highest grades, despite a higher proportion of boys taking these subjects. This gap has now disappeared. So, the

follow up questions include why the gaps appear only at high levels of achievement, and why they differ over time and between subjects? Is it changes in subject entry, the nature of the teaching, or the form of assessment? Once again, more evidence, more data, is needed. The initial secondary analysis helps the study onto the most appropriate track but it represents a starting point only.

Table 1 - Achievement gap in favour of girls at each grade, A level Mathematics, Wales, 1992-1997

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1992 | -7 | -4 | -2 | -1 | 1 | 1 |
| 1993 | -15 | -9 | -5 | -3 | -1 | -1 |
| 1994 | -5 | -2 | 0 | 0 | 0 | 0 |
| 1995 | -3 | -2 | -2 | 0 | 0 | 0 |
| 1996 | -5 | 2 | -2 | 0 | 0 | 0 |
| 1997 | 0 | 1 | 0 | 0 | 0 | 0 |

Note: The achievement gap for any grade is calculated as the number of girls attaining that grade, minus the number of boys attaining that grade, all divided by the total attaining that grade.

Secondary data can provide the evidence for an important stand-alone initial analysis. For example, to find out whether the number of applicants to study undergraduate science at universities in the UK has been going up or down in the last 10 years it is difficult to imagine collecting better data on this than the Higher Education Statistics Agency (HESA) already does. It is possible to conduct a new analysis of these data and produce original publishable research that could be important in terms of policy and practice (e.g. Gorard 2008a). HESA also conduct an annual destination survey, which gathers information on the activities of graduates six months after they graduate. This dataset only considers destinations at six months after the student has left university, and career trajectories may be very different subsequently. Nevertheless it remains the most complete picture of the first occupation of graduates.

Using these figures, Smith and Gorard (2011) recently posed the question, is there actually any shortage of scientists in the UK? Their answer was that STEM graduates are unlikely to remain in the field after university – other than as teachers of others. An example snapshot of their analysis is that every year, for well over 50% of engineering graduates, their first occupational destination is not directly related to engineering, or is unknown. Maybe, because of recent initiatives, there are too many people studying engineering for the labour market to cope with, or perhaps graduates are no longer of sufficient quality for employers. But it is perhaps more likely that so many engineers are without relevant employment every year because the shortage thesis is wrong and there are no jobs waiting for all of them, or because they are 'dropping out' having learnt that they do not enjoy their subject area. It is not possible to decide between these alternatives using the secondary data presented. Secondary analysis, like most research, leads naturally to further questions and study. Use of a large scale dataset is frequently the start of further investigation, not an end in itself. And interestingly, just as with earlier political arithmetic, laying bare a pattern like this often leads to initial disbelief, then outrage, and even a view that it would have been better to suppress the data pattern (e.g. "It's nonsense to claim Britain produces too many science graduates", The Guardian, 14/9/11).

Extremely large-scale, long-term and official datasets also carry a certain authority, and this can be reflected in any further work involving the same data. A dataset like the Labour Force Survey (LFS), covers hundreds of variables relating to 150,000 people collected every three months, with the results available since 1973. Whatever its faults, it is clearly of a much higher quality than anything most researchers could ever hope to achieve in a small project. A claim that job-related training for over-35s has declined in the UK over the last 10 years is more likely to be believed (and quite rightly so) if the evidence is a re-analysis of the LFS than if it is a survey of 100 people. Yet, it will be both quicker and easier to use the LFS data than to collect 100 survey responses. Like the other datasets listed below, the LFS data comes with exemplary documentation and advice

**Some sources of secondary data**

Much of the existing data that might be useful for education and social science researchers is available for download from websites, or can be requested from official bodies. Some likely sources are suggested here for illustration, predominantly from the UK, but the details of internet resources are likely to date rapidly, and will vary between countries.

The (Office for) National Statistics is a one-stop shop for evidence on almost anything (http://www.ons.gov.uk/ons/index.html). It offers figures on agriculture, fishing and forestry, commerce energy and industry, health and care, the labour market, the natural and built environment, population and migration, welfare, transport travel and tourism, the time-use survey, and several other themes. It includes evidence at small area level on all ten-yearly national censuses of the population, most recently from 2011. The UK population Census contains questions about type of housing, car ownership, household relationships, economic activity, health, and qualifications. Many of the questions have been asked every ten years since 1841. ONS also has a special section with datasets on children, education and skills, with a wealth of data, including on regional trends, pupil:teacher ratios for all schools, and an Atlas of Deprivation. The Social Focus on Children is a summary of UK statistics relating to children, such as what they read, how they spend their money, and what their leisure interests are. Social Trends is an annual production since 1970 giving figures on education, health, employment, leisure, transport and housing. Family Spending reports the findings of the regular Family Expenditure Survey showing how households distribute their incomes between food, travel, housing and other demands. Statistics of Education UK shows the annual figures for many education related topics (with past years to 1972 for comparison) including the number of teachers and students by school and sector, and participation and qualification rates for each age group of students. ONS also publish descriptions of public policy systems, such as education, in other countries, and annual reports of the first destinations of UK graduates, and trends and predictions for the supply of graduates to industry. Other key themes are crime, justice, offenders, and terrorism. These are dealt with in the British Crime Survey, with data on crime rates, fear of crime, recorded crime, attitudes to crime, and crime reduction. These are all broken down by area and type of crime, and allow the examination of trends over time.

The national UK Data Archive (http://www.data-archive.ac.uk/home) is a repository of all datasets generated through research paid for by the taxpayer-funded Research Councils (such as the Economic and Social Research Council), and from a number of other sources. It includes historical archives, policy and other documents, and transcripts of interviews undertaken as part of previous research projects. It includes large-scale surveys on adult literacy, patterns of lifelong learning, and the British Household Panel Survey. It also provides access to, international datasets including such diverse sources as Bulgarian microdata, US marital instability over the lifecourse, UNESCO Education Database, the Dutch Panel Survey, and even the physical stature of Georgia convicts from 1770-1860, for example. See also the National Digital Archive of Datasets (http://www.nationalarchives.gov.uk.ezproxye.bham.ac.uk/documentsonline/datasets.asp), including a database of the annual schools census for all schools in England, with data at school level on pupil intake characteristics (poverty, special needs, ethnicity, sex, first language) and on the teaching and support staff. The Department for Education has a website full of data on all aspects of school and childhood, including an archive of examination and key stage results for each school up to the current year (http://www.education.gov.uk/performancetables/). Edubase is a set of data about every educational institution in England, summarising their intake, management, whether they are in special measures (for schools) and even including information on the population density of the locale. The publicly available component can be accessed at http://www.education.gov.uk/edubase/home.xhtml. Many other official and government websites include downloadable data. One of these is the Higher Education Statistics Agency (http://www.hesa.ac.uk/). This has an archive of applications and admissions to higher education, broken down by region, subject and institutions among other things. Other UK and England-based data can be found via the Training and Development Agency, the Office for Standards in Education, the Basic Skills Agency, and many other bodies.

Beyond the UK, the OECD website has a collection of international educational evidence, including the annual Education at a Glance which has sections on work-based and tertiary education as well as schooling. It also contains the results of successive rounds of the international PISA study (http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html). The most recent PISA study was in 2009, and includes the views of teachers and students, student test results in a range of subjects, and school-level data. It can be downloaded from the website, giving records for individuals within schools, in around 80 countries. Most developed countries have equivalent sources of national data. Some have longitudinal studies, such as the US High School and Beyond Survey, and the US Cohort Study similar to the UK National Child Development Study.. The European Union produces a variety of statistical summaries allowing comparisons of most European public policy systems, and the socio-economic systems from which they emerge (through CERI and Eurostat).

This summary can only give a glimpse of the huge and growing number of sources of high quality data relevant to education and social science. For almost all research topics, the difficulty is not so much whether some relevant data exists but how rapidly it improves in range, completeness, quality and access. Some of the website links above will probably change between writing this paper and publication. The data will still be there, and in even greater quantity, but the form of access may have altered.

## The limitations of secondary data

There are, of course, limitations in using large-scale datasets collected by others for another purpose. The data may be less than ideal, and the conditions under which the data was collected may be unclear or relatively undocumented, sometimes missing the field notes or other incidental observations made during the process of primary collection. The equivalents have to be provided otherwise. When looking at improvement in England in scores for Key Stage assessments in schools since their inception, it is important to recall the early disruption caused by teachers' lack of cooperation. When using examination figures in any way it is important to recall the difficulties of ensuring comparability of standards between subjects, syllabuses, examination boards, years, modes of assessment, and regions. A secondary analysis, done well, requires a thorough examination of the pedigree of its raw materials. But none of these limitations are usually easier for, or improved by, small-scale primary data collection. The same concerns about purpose, completeness, lack of bias, intervening events, and so on, should occur to the reader of an account of the primary research of others. For all research, we should be aware of the limitations of the data we are using. We should publish these limitations, and take them into account in our findings (Gorard 2008a). Otherwise the danger is that, as in the recent example of the purportedly declining social mobility in the UK, the authority of large datasets can induce research users to waste time and resources (Gorard 2008b). Secondary data are useful only insofar as their quality and analysis warrants.

## How are large-datasets analysed?

One of the growing concerns about large existing datasets is about how to analyse them. We have few generally agreed methods for dealing with these often complex, secondary datasets, which means that many of those researchers using secondary data have tended to be pioneers of one kind of analysis or another. There are currently debates between them over the precise way to measure trends over time, differences between places, and how to deal with hierarchical data for example.

Overall, there seem to be two somewhat divergent trends. One version continues the political arithmetic tradition of laying bare the patterns in the data for the widest possible audience using the simplest of techniques. The justification for this is both ethical/democratic and methodological. Since so many official datasets are for populations in the statistical sense (including the cohort studies of 1958, 1970 and 2000, the national pupil database, and university entry records, for example), no statistical generalisation is needed or possible. Many other datasets, such as the Labour Force Survey and other panel approaches, are not based on random samples. If, using the National Pupil Database, female pupils attained a higher average GCSE points score than males in England in 2011, then that is the end of the matter. An analyst can calculate and present that difference, or comment on it. They cannot and should not ask if that difference is statistically significant, or whether there ‘really’ is a difference (Gorard 2010a). The commentary might consider how large the difference is compared to other years, other phases, other countries, or how large it is in relation to what is known about missing data and errors in the measurement

process. None of these issues relates to random sampling and so no statistics, as traditionally conceived, is involved.

Even for other datasets the analytical concern is not so much with random sampling variation as with how to account for erroneous and missing data. Therefore, the panoply of sampling theory statistics from chi-squared tests to MANCOVA is largely irrelevant. The reporting of significant differences between cases in population data, or quoting confidence intervals for UK higher education league tables, for example, are simply statistical misunderstandings (see also Gorard 2010b). Shorn of these, secondary data analysis requires almost pedantic care and attention, but it is not complex either to conduct or to describe to research users. The job of the analyst is to make clear any relevant patterns in the data so that they receive wider inspection.

Of course, large datasets of population figures can be also modelled using regression techniques and similar, and the results can be fascinating. But such modelling is not essential and does not represent any kind of definitive test. Mostly, large datasets can be analysed as simple frequencies and/or percentages, broken down into categories such as year, sex of student, or geographic region. It is slightly more complicated when cross-analysing two or more large datasets, but even here the complication relates to the organisation of the datasets rather than the technique of analysis as such.

The alternate trend in analysis is towards the exact opposite. Some analysts claim that even population data is really only a sample from an imagined 'super-population'. So, they defend the use of sampling theory derivatives such as significance testing with almost all datasets. But this leaves them with a problem because the existing datasets are often complex in structure, with natural hierarchical clusters, such as pupils within classes within schools. This, they claim, then necessitates the use of especially complex methods of analysis to compensate for theoretical deficiencies in the data. Consequently and unfortunately, work of this kind no longer has the appeal of laying bare the body politic, since so few people can currently understand the research or its findings – and it is largely for this reason of un-readability that no examples are included here. Whether probabilistic approaches are valid for non-probabilistic data, whether statistical analysis can really do anything about erroneous and missing data, and whether complex data genuinely always need complex methods for analysis, are all issues increasingly covered in BJES and elsewhere over the last 15 years (Gorard 2003). For the present, it seems that the two trends will continue in isolation for the most part. The next section looks at two sub-fields of research that exemplify the two trends in analysis.

## Origins and destinations and school effectiveness in the UK

In the UK, a seminal piece of work in the political arithmetic tradition was 'Origins and Destinations' by Halsey et al. (1980). By relating individual learner characteristics to their schooling, it showed the importance and stability of family and social class in educational and labour outcomes. The analysis was largely the usual laying bare the patterns of inequality. What is surprising on re-reading it today is how poor were the datasets the authors had to contend with, compared to what might be available now. This renewal of political arithmetic was continued by the Nuffield Sociology Group (Marshall et al. 1997, Heath 2000), and by its 'offshoots' (Sullivan

et al. 2011). Much valuable related work has been based on longitudinal birth cohort studies (Breen and Goldthorpe 2001, Hobcraft 2001, Bynner and Joshi 2002). And concern about class origin has extended to consideration of sex, ethnicity, learning difficulties and many other potentially stratifying variables for school and lifelong outcomes (Gorard et al. 2001, Gorard and Selwyn 2005). The overall approach has been largely un-contentious, with academic disputes mostly concerned with the precise ways in which absolute and relative differences should be presented (Gorard 1999, 2006).

In contrast, school effectiveness is a field of research, started in the US and developed in the UK from around the time of 'Fifteen Thousand Hours' by Rutter et al. (1979). Much recent educational research now stems from the influence of this school effectiveness movement. The field has attempted to identify differentially successful schools, and to describe their characteristics in a way that could form the basis of a blueprint for wider school improvement. The work has been continued by a number of analysts, such as Gray et al. (1986) and Nuttall et al. (1989), and in the US has been extended to consideration of the effectiveness of individual teachers (Sanders 2000). As with the 'origins' approach, the field of school effectiveness has benefited from the increasing quality of data, especially about individual pupils. In addition, it has been associated with the development of a range of complex methods of analysis – most notably hierarchical linear modelling (Aitkin and Longford 1986). The challenges to the school effectiveness approach have tended to be more fundamental than to political arithmetic, but not terribly successful (Slee et al. 1998), although some more technical commentaries have been more important (Amrein-Beardsley 2008).

Intriguingly, the major undisputed outcome of both streams of work has been the reinforcement of the importance of individual pupil characteristics in predicting school and later educational outcomes. Early large-scale studies by Coleman et al. (1966) and Jencks et al. (1972) in the US showed that once individual student characteristics had been taken into account, very little of the variation in school examination outcomes was left to be explained by a 'school effect'. It appeared that any individual would be expected to attain pretty much the same results whichever school they attended. Knowing each pupil's level of attainment prior to any phase of schooling, and as much as possible about their socio-economic and family background permits a very accurate prediction of each pupil's level of attainment *after* that phase of schooling. Since 1966, the quality and range of individual-level data has improved, prior attainment has been factored in, and the predictions are even more accurate than before. As the data and models built on them improve over time, and as the number of cases grows, so the strength of the relationship between pupil characteristics and their educational outcomes grows. This leaves less and less residual variation to be explained by any combination of error terms, and institutional and other educational processes.

What has largely differed between the two traditions then has been the reaction of researchers to this central finding, and the practical importance attached to their results by practitioners and policy-makers. Bynner and Joshi (2002) drew on two successive longitudinal cohort studies that commenced in 1958 and 1970 respectively, constituting perhaps the most powerful UK datasets linking social background and modes of schooling to educational achievement. These demonstrate the extent to which social class remained the strongest determinant of student attainment and school performance over time. Social class, rather than, for example, systemic change from selective to

comprehensive education emerged as the greatest influence on the age at which students left school, their leaving qualifications and their chances of entering higher education. 'Over the past 25 years... studies show that individual and family background traits explain the vast majority of the variance in student test scores, and observable school characteristics, such as per student spending, teacher experience, or teacher degree level, have at best a weak relationship with student outcomes' (Goldhaber et al. 1999, p.199). Many large studies reach the same conclusion (Suter 2000). The message often taken is that schools are not over-coming initial pupil disadvantage, and so education seems to be a very poor from of social engineering. It seems that initial disadvantage must be addressed directly (Gorard et al. 2011).

In the school effectiveness tradition, however, Reynolds (1990) stated that there were 'large school effects' (p.154), because up to 8% of the variance in school results was due to school effects rather than individual characteristics. Other studies have repeatedly shown the same kind of scale for the proposed school effect – anywhere from zero (Shipman 1997), through 8% to 10% (Daly 1991), to 12% and possibly a bit more (Creemers 1994, Stoll and Fink 1996). In a review of effectiveness studies, Gray and Wilcox (1995) attributed from 2% to 10% to school effects. However, rather than concluding that around 90% or more of the variation in school outcomes was due to prior individual characteristics, the focus of these commentators has been on the smaller component that could be school effect (or could be error). These writers are saying, in contradiction to the earlier studies from the US, that it matters which school a pupil attends, and that there are 'good' schools and 'bad' schools. Converted to examination grades or points, the 10% variation left unaccounted for by pupil background could make a key difference to a pupil's life chances - but only if it truly is a school or teacher effect.

Due to such claims, adherents of school effectiveness have been influential in education policy in the UK and elsewhere. Barber (1996), for example, concluded from a review of the evidence that 'the research into school effectiveness over the last two decades has made an immense contribution to our understanding of school performance. Whereas in the 1960s and early 1970s the prevailing view of education and social researchers was that the effect of school on a pupil's performance was negligible in comparison to the impact of social class and upbringing, it is now demonstrable that schools make a significant difference to how well children do' (p.127). Barber was an academic who was then recruited to oversee the Standards and Effectiveness Unit by the New Labour government elected in 1997. He now offers advice to supra-national governments on school and teacher effectiveness (Barber and Mourshed 2007). School effectiveness has become part of official discourse. The attraction for policy-makers is that education – or indeed 'education, education, and education' – can be made responsible for overcoming inequality in society and then blamed if it fails to do so. Schooling seems like a malleable process open to ameliorative intervention, in a way that sex, ethnicity and even social class are not.

There are two problems with this approach. First, it is not clear to all that the 10% unexplained by pupil origin is a genuine and lasting school 'effect'. It is very volatile across years, and for different kinds of outcomes (Sammons et al. 1996). Second, the recipes for ameliorative intervention or school improvement do not seem to work. When researchers have attempted to relate the small school-effect to school characteristics and processes, so producing a blueprint for school improvement, the

results have generally been negligible. The factors making up a 'good' school are frequently rather nebulous or blindingly obvious and tautological, consisting of items like an academic emphasis or high-quality leadership or good discipline (Ouston 1998, Gorard 2000). Even the school impact that has been recorded cannot be seen as a unitary trait applying to all subjects, departments, ages, and abilities, and to both sexes (Nuttall et al. 1989). In fact, the worry is that it is easy to begin to imagine that schools make no difference at all (Rothstein 2002). Of course, this is wrong. Teachers are not wasting their time. The question is not whether schools make a difference, in comparison to not having schools, for example. It is not even whether schools, as currently envisaged are the most efficient way for society to achieve whatever it is that it wants from school. School effectiveness does not address any of these. It focuses solely on whether any school is noticeably more or less effective in terms of pupil attainment than any other.

The irony is that when school effectiveness models based on explaining the residual variance are used to push for small-scale policy changes like homework clubs or compulsory uniforms, their findings are often seized upon by ministers. When nearly identical methods are used to point out the dominant role of socio-economic context in school outcomes, then politicians appear to be less happy, and worry about accounting for social background when assessing performance, in case they are accused of fiddling the figures or obscuring the real level of performance. 'The problem of underperformance has been largely... conceived of as a failure of schools and of teachers... What School Effectiveness Research has failed to provide... is to develop an understanding of the processes which have led to the remarkably strong and surprisingly consistent relationship between socio-economic context and school performance' (Gibson and Asthana 1998, p.207).

## Conclusion

If current trends continue, then the use of secondary data will become even more important for education, both in research and policy-making. The range and quality of datasets will improve even further. There will be imaginative cross-pollination of two or more seemingly unrelated datasets to create occasional paradigm shifts or new sub-disciplines, like that caused by ‘Origins and Destinations’ or by the pioneers of school effectiveness. Perhaps researchers will realise that examining the existing data is at least as important as examining existing accounts of evidence when synthesising materials at the start of a new project. The electronic databases of abstracts and reports from PsychInfo to Sociological Abstracts certainly mean that there is no excuse in the future, where literature review is the intention, for such partial and limited reviews of the literature as have been produced by so many authors in the past (often no more than citation clubs in fact).

In addition, I feel that the two basic approaches to secondary analysis will diverge further. One, stemming more from the sociology of education, may use the opportunity of open government, freedom of information and web-based resources to lay bare more injustice and incompetence, using relatively simple and inclusive forms of analysis. The other approach, stemming more from psychology, may continue to focus on matching individuals more and more carefully, in order to ‘squeeze’ out the identifiable malleable factors for improving education. Both fields have been

valuable, and either may continue to provide insights. But it is time, I feel, for both to move on, further into the cycle of a programme of research (Gorard 2010c). The rest of this paper suggests why.

In this paper, the first approach has been exemplified by work which emphasised the unfairness of education, but could be accused of assuming sometimes that this alone would be embarrassing enough to provoke an appropriate policy or practice response. To a large extent this has not worked. Perhaps the shock has dissipated. In a sense, everyone knows about the structural inequalities in education but no one with influence in education or politics really knows what to do about them. Perhaps what education needs to do is stop colluding with policy-makers in imagining that education can solve most social ills. Education could hand back the problem and the kudos that came with the pretence that the problem could be solved. Or more usefully, the field needs to move on from its simple laying bare approach and begin to find the underlying malleable causes of the manifested educational inequalities. There are indications of promising interventions to overcome these inequalities (Gorard et al. 2011). Perhaps it is time to embrace the development of solutions rather more, through rigorously evaluated intervention studies – moving from what the US government Institute of Education Science (IES - ies.ed.gov/funding/webinars/previous_webinars.asp) would term goal one and two work (exploratory and development) to goal three and four (efficacy trials and scaling up). It is time for political arithmetic to grow up.

The second approach uses similar premises and objectives, relating school outcomes to individual prior characteristics. However, in school effectiveness work the concern is not so much with laying bare the social stratification of outcomes but with attributing any residual variation to the impact of schools (or departments or teachers, or whatever). The idea was that by identifying the nature of especially effective schools, all schools could benefit and those pupils facing the weakest outcomes might benefit the most. School effectiveness has emphasised the determinants of the outcomes that were malleable, and so contained the seeds of its own prescriptions for educational improvement. These prescriptions have not been tested in any rigorous way, and there is therefore a danger of falling for the superstitious interpretation of correlations that Ouston (1998) calls the potted plant theory. I feel that the field also needs to move on in the same way as political arithmetic. There has been enough dabbling in goal one and two work. It is time either to admit that nothing of value has been developed by school effectiveness and cease this work, or to make clear and testable claims for school improvement that can be causally evaluated by rigorous interventions. To do neither is both wasteful and unethical.

Secondary data analysis is important and valuable but it is not complete in itself. It cannot identify causal models of the kind that school and societal improvement need. Whereas the ‘origins’ approach relies heavily on a limited range of research designs, especially longitudinal and historical approaches, and analyses data accordingly, the effectiveness approach generally eschews design and utilises increasingly complex statistical approaches with snapshot datasets. The next obvious step in both fields is not more data dredging and not more complexity, but the identification of likely interventions for educational improvement, and their development and evaluation through randomised controlled trials (or similar). The two fields could even come together to achieve this, although I suspect they will not.

## References


Aitkin, M. and Longford, N. (1986) Statistical modelling in school effectiveness studies, *Journal of the Royal Statistical Society Series A*, 149, 3, 1-43

Amrein-Beardsley, A. (2008) Methodological concerns about the education value-added assessment system, *Educational Researcher*, 37, 2, 65-75

Barber, M. (1996) *The learning game*, London: Indigo

Barber, M. and Mourshed, M. (2007) *How the world's best-performing school systems come out on top*, McKinsey & Co.

Breen, R. and Goldthorpe, J. (2001) Class, mobility and merit: the experience of two British birth cohorts, *European Sociological Review*, 17, 81-101

Bulmer, M. (1980) Why don't sociologists make more use of official statistics?, *Sociology*, 14, 4, 505-523

Bynner J. and Joshi H. (2002) Equality and opportunity in education: the evidence from the 1958 and the 1970 birth cohort surveys, *Oxford Review of Education*, 28,4, 405-425.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F. and York, R. (1966) *Equality of educational opportunity*, Washington: US Government Printing Office

Creemers, B. (1994) The history, value and purpose of school effectiveness studies, in D. Reynolds, B. Creemers, P. Nesselradt, E. Shaffer, S. Stringfield and C. Teddlie (Eds.) *Advances in school effectiveness research and practice*, Oxford: Pergamon

Daly, P. (1991) How large are secondary school effects in Northern Ireland?, *School Effectiveness and School Improvement*, 2, 4, 305-323

Gibson, A. and Asthana, S. (1998) School performance, school effectiveness and the 1997 White Paper, *Oxford Review of Education*, 24, 2, 195-210

Goldhaber, D., Brewer, D. and Anderson, D. (1999) A three-way error component analysis of educational productivity, *Education Economics*, 7, 3, 199-208

Gorard, S. (1999) Keeping a sense of proportion: the "politician's error" in analysing school outcomes, *British Journal of Educational Studies*, 47, 3, 235-246

Gorard, S. (2000) *Education and Social Justice*, Cardiff: University of Wales Press

Gorard, S. (2001) Quantitative Methods in Educational Research: The role of numbers made easy, London: Continuum

Gorard, S. (2002) Political control: A way forward for educational research?, *British Journal of Educational Studies*, 50, 3, 378-389

Gorard, S. (2003) In defence of a middle way: a reply to Plewis and Fielding, *British Journal of Educational Studies*, 51, 4, 420-426

Gorard, S. (2006) Yet another perspective: a response to Connolly, *British Journal of Educational Studies*, 54, 4, 471-475

Gorard, S. (2008a) Who is missing from higher education?, *Cambridge Journal of Education*, 38, 3, 421-437

Gorard, S. (2008b) Research impact is not always a good thing: a re-consideration of rates of 'social mobility' in Britain, *British Journal of Sociology of Education*, 29, 3, 317-324

Gorard, S. (2010a) All evidence is equal: the flaw in statistical reasoning, *Oxford Review of Education*, 36, 1, 63-77

Gorard, S. (2010b) Serious doubts about school effectiveness, *British Educational Research Journal*, 36, 5, 735-766
Gorard, S. (2010c) Research design, as independent of methods, in Teddlie, C. and Tashakkori, A. (Eds.) *Handbook of Mixed Methods*, Los Angeles: Sage
Gorard, S. and Fitz, J. (1998) The more things change.... the missing impact of marketisation, *British Journal of Sociology of Education*, 19, 3, 365-376
Gorard, S. and Rees, G. (2002) *Creating a learning society?*, Bristol: Policy Press
Gorard, S. and Selwyn, N. (2005) What makes a lifelong learner?, *Teachers College Record*, 107, 6, 1193-1216
Gorard, S. with Taylor, C. (2004) *Combining methods in educational and social research*, London: Open University Press
Gorard, S., Rees, G. and Salisbury, J. (2001) The differential attainment of boys and girls at school: investigating the patterns and their determinants, *British Educational Research Journal* , 27, 2, 125-139
Gorard, S., See, BH and Davies, P. (2011) *Do attitudes and aspirations matter in education?: A review of the research evidence*, Saarbrucken: Lambert Academic Publishing
Gorard, S., Taylor, C. and Fitz, J. (2003) *Schools, Markets and Choice Policies*, London: RoutledgeFalmer
Gray, J. and Wilcox, B. (1995) *'Good school, bas school' Evaluating performance and encouraging improvement*, Buckingham: Open University Press
Gray, J. Jesson, D. and Jones, B. (1986) The search for a fairer way of comparing schools' examination results, *Research Papers in Education*, 1, 2, 91–122
Hakim C. (1982) Secondary analysis in social research: a guide to data sources and methods with examples, London: Allen and Unwin
Halsey, A., Heath, A. and Ridge, J. (1980) *Origins and Destinations: Family, Class, and Education in Modern Britain*, Oxford: Clarendon
Heath, A. (2000) 'The political arithmetic tradition in the sociology of education', *Oxford Review of Education* 26: 313-331.
Hobcraft, J. (2001) *Intergenerational transmission of inequality in a British birth cohort*, presented at the Population Association of America Annual Meeting, Washington DC, 29-31st March 2001
Jencks, C., Smith, M., Ackland, H., Bane, M., Cohen, D., Gintis, H., Heyns, B. and Nicholson, S. (1972) *Inequality: assessment of the effect of family and schooling in America*, New York: Basic Books
Marshall, G., Swift, A. and Roberts, S. (1997) *Against the odds? Social class and social justice in industrial societies*, Oxford: Clarendon Press
Nuttall, D., Goldstein, H., Presser, R. and Rasbash, H. (1989) Differential school effectiveness, *International Journal of Educational Research*, 13, 7, 769-776
Ouston, J. (1998) *The school effectiveness and improvement movement: a reflection on its contribution to the development of good schools*, presented at ESRC Redefining Education Management seminar, Open University, 4/6/98
Porter, T. (1986) *The rise of statistical thinking*, Princeton: Princeton University Press
Rendall, M. (2003) *Quantitative research: a scoping study*, London: Learning and Skills Research Centre
Reynolds, D. (1990) School effectiveness and school improvement: a review of the British literature, in Moon, B., Isaac, J. and Powney, J. (Eds.) *Judging standards and effectiveness in education*, London: Hodder and Stoughton
Rothstein, R. (2002) *Out of balance: our understanding of how schools affect society and how society affects schools*, Chicago: The Spencer Foundation

Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours: Secondary schools and their effects on children*. London: Open Books.

Sammons, P. Mortimore, P. and Thomas, S. (1996) Do schools perform consistently across outcomes and areas?, pp. 3-27 in Gray, J., Reynolds, D., Fitz-Gibbon, C. and Jesson, D. (Eds.) *Merging traditions: the future of research on school effectiveness and school improvement*, London: Cassell

Sanders, W. (2000) Value-added assessment from student achievement data, , *Journal of Personnel Evaluation in Education*, 14, 4, 329-339

Selwyn, N., Gorard, S. and Furlong, J. (2006) *Adult learning in the digital age*, London: RoutledgeFalmer

Shipman, M. (1997) *The limitations of social research*, Harlow: Longman

Slee, R., Weiner, G. and Tomlinson, S. (1998) *School Effectiveness for Whom? Challenges to the school effectiveness and school improvement movements*, London: Falmer Press

Smith, E. (2008) Using secondary data in educational and social research, Maidenhead: Open University Press

Smith, E. and Gorard, S. (2011) Is there a shortage of scientists? A re-analysis of supply for the UK, *British Journal of Educational Studies*, 59, 2, 159-177

Stoll, L. and Fink, L. (1996) *Changing our schools: linking school effectiveness and school improvement*, Buckingham: Open University Press

Sullivan, A., Joshi, H. and Leonard, D. (2011) Single-sex schooling and labour market outcomes, Oxford Review of Education, 37, 3, 311-332

Suter, L. (2000) Is student achievement immutable?, *Review of Educational Research*, 70, 4, 529-545